Remarks

Claims 1-11 are pending in the application. Claims 1-11 are rejected. Claims 1, 3 and 11 are amended. Claim 6 is canceled. All rejections are respectfully traversed.

Claim 1 is amended to incorporate the limitations of claim 6. Claim 3 is amended to depend from claim 2. Claim 6 is canceled. Claim 11 is amended to correct a typographical error.

The invention summarizes a compressed video. A cumulative motion activity intensity is measured from an average motion vector magnitude in the compressed video. Key frames are selected from the compressed video according to the cumulative motion activity intensity and the key-frames are concatenated in a temporal order to form a summary of the compressed video.

Claims 1-11 are rejected under 35 U.S.C. 102(e) as being anticipated by Toklu et al. (U.S. Patent No. 6,549,643).

Toklu summarizes an uncompressed video by partitioning the video into segments and generating a temporal activity curve for dissimilarity measures of each segment based on frame differences, color histograms, and/or camera motion.

The invention measures a cumulative motion activity intensity from an average motion vector magnitude in a compressed video.

The Examiner cites Toklu, column 9, lines 17-46, as teaching the claimed measuring. However, it is clear from Toklu, column 9, lines 30-37 (reproduced below), that the method of Toklu is pixel-based and, therefore, operates on uncompressed video.

> where N denotes the total number of pixels in every frame.
>
> The temporal MA curve is then reduced to a Binary MA curve by comparing the MA curve to predefined threshold on a frame-by-frame basis (step **214**) to detect frames having significant motion (motion activity segments within the shot). More specifically, for each frame in a given shot, the MA curve is compared to a predefined threshold that is preferably set to 0.5 pixels. The computed MA value for a

Those of ordinary skill in the arts would understand that compressed videos do not have pixels. Instead, compressed videos include and operate on DCT coefficients and motion vectors. Operations on pixels of uncompressed videos as described in Toklu cannot anticipate operations on compressed videos as claimed. Toklu is non-analogous art.

Toklu estimates camera motion from dissimilarity measures obtained by comparing images in uncompressed video on a frame-by-frame basis. Toklu cannot teach the claimed measuring using average motion vector magnitudes, which are only available in the compressed domain.

Also claimed is selecting key-frames from the compressed video according to the cumulative motion activity intensity. As was shown above, Toklu does not measure a cumulative motion activity intensity from an average motion vector magnitude in the compressed video. Toklu cannot teach the claimed selecting of key-frames according to the cumulative motion vector activity intensity.

Also claimed is concatenating the key-frames in a temporal order to form a summary of the compressed video. The Examiner cites Toklu, column 6, lines 35-41 (reproduced below), as teaching the claimed concatenating.

> The output of the key-frame elimination module **18** is list of key-frames, e.g., keyframelist, for each video shot. The key-frame list may be stored in a key-frame list database **19**. The key-frame lists may be used by video processing/ indexing systems as components for, e.g., multimedia and/or video summarization.

Toklu makes no mention of a temporal order for the list of key-frames. Further, the invention concatenates key-frames in compressed video. Toklu's pixel-based method does not operate in the compressed domain and cannot teach the claimed concatenating.

Operations on uncompressed videos, as described by Toklu, cannot anticipate operations on compressed videos. Furthermore, camera motion while acquiring an uncompressed video, see

> by-frame basis. Preferably, the camera motion is estimated by estimating motion parameters such as zooming, rotational, panning, and tilting. The key-frames of a given

has nothing to do with motion activity in a compressed video. Anyone of ordinary skill in the art would know that motion vectors as claimed are only available in compressed videos.

Regarding claim 2, claimed is partitioning the compressed video into a plurality of segments, each segment having a substantially equal amount of cumulative motion activity intensity, and selecting one key frame from each segment. The Examiner cites Toklu, column 5, lines 50-60 (reproduced below), as teaching the claimed partitioning.

> issued on Nov. 10, 1998, entitled "Apparatus For Detecting A Cut In A Video," which is commonly owned and incorporated herein by reference. The above-incorporated cut detection method partitions video data into a set of "shots" comprising visually abrupt cuts or camera breaks. As stated above, a video "shot" represents a contiguous recording of one or more video frames depicting a continuous action in time and space. The above-incorporated cut detection method segments video data into shots using an integrated pixel-based difference metric (discussed below) and distribution-based difference metric.

The method of Toklu can never anticipate the claimed partitioning because it operates on pixels of uncompressed video and creates "shots" bounded by "visually abrupt cuts or camera breaks", not segments having a substantially equal cumulative motion activity as determined by measuring average motion vector magnitude in compressed video. The Examiner also cites Toklu, column 10, lines 5-15, 6-50 and 51-63 (in sum, lines 5-63), as teaching the claimed partitioning and selecting. In that section, it is clear that Toklu uses threshold criteria for selecting key-frames, see Toklu, column 10, lines 35-45 (reproduced below), and that the motion in Toklu is camera motion (panning and zooming), which have nothing to do with motion vectors.

> **219)**. When either the (1) cumulative sum for the panning motion exceeds a first predefined threshold (which is preferably two-thirds of the frame width) (affirmative result in step **220**), (2) cumulative sum for the tilting motion exceeds a second predefined threshold (which is preferably two-thirds of the frame height) (affirmative result in step **221**), or (3) the cumulative sum of the absolute value of the zooming motion exceeds a third predefined threshold (which is preferably 0.66) (affirmative result in step **222**), then the current frame is selected as a key-frame for that motion activity segment, all the cumulative sums are reset to zero, and the

However, as was shown above, the claimed invention partitions compressed video and selects key-frames according to cumulative motion activity intensity. Toklu cannot teach the claimed partitioning and selecting.

Regarding claim 3, claimed is partitioning a compressed video into a number of segments one less than a number of desired key-frames in the summary. The Examiner cites Toklu, column 10, lines 40-45 (reproduced below), as teaching the claimed partitioning.

> thirds of the frame height) (affirmative result in step **221**), or (3) the cumulative sum of the absolute value of the zooming motion exceeds a third predefined threshold (which is preferably 0.66) (affirmative result in step **222**), then the current frame is selected as a key-frame for that motion activity segment, all the cumulative sums are reset to zero, and the

Applicant cannot find any mention of either numbers of segments or numbers of desired key-frames in the summary in the cited section. Toklu does not teach the claimed partitioning into a number of segments one less than a desired number of key-frames in the summary.

Regarding claim 4, claimed is selecting a first frame of each segment as one of the key-frames of the summary and selecting a last frame of the compressed video as a last key-frame of the summary. As was shown above, Toklu partitions uncompressed video into "shots" bounded by "visually abrupt cuts or camera breaks," not segments having a substantially equal cumulative motion activity as determined by measuring average motion vector magnitude in compressed video, as claimed.

Also, the method of Toklu selects key-frames using threshold criteria, not according to cumulative motion activity intensity, as claimed. Furthermore, Applicant can find no specific rationale in the Examiner's action for the rejection of claim 4. Therefore, Applicant regards claim 4 as allowable.

Regarding claim 5, claimed is selecting a middle frame according to the accumulated motion activity intensity of each segment as one of the key-frames. As above, Toklu does not select key-frames according to cumulative motion activity intensity and, as a result, cannot teach the claimed selecting. Furthermore, Applicant can find no specific rationale in the Examiner's action for the rejection of claim 5. Therefore, Applicant regards claim 5 as allowable.

Regarding claim 7, claimed is measuring the motion activity intensity from a median motion vector magnitude. The Examiner cites Toklu, column 9, lines 53-57 (reproduced below), as teaching the claimed measuring.

> smoothing process is to eliminate the holes that are less than or equal to 3 frames long. Next, the binary MA curve is median filtered using a 7-tab filter (as is understood by those skilled in the art). The next step is to eliminate holes that are less than or equal to 15 frames long and finally, to eliminate

Those skilled in the art would never confuse a median motion vector magnitude with a process for median-filtering a curve. Median-filtering is used to remove noise from signal. This cannot teach the claimed measuring from a median motion vector magnitude.

Regarding claim 8, claimed is measuring motion activity intensity from a standard deviation of the motion vector magnitude. The Examiner cites Toklu, column 10, lines 35-45 (reproduced below), as teaching the claimed measuring.

> **219**). When either the (1) cumulative sum for the panning motion exceeds a first predefined threshold (which is preferably two-thirds of the frame width) (affirmative result in step **220**), (2) cumulative sum for the tilting motion exceeds a second predefined threshold (which is preferably two-thirds of the frame height) (affirmative result in step **221**), or (3) the cumulative sum of the absolute value of the zooming motion exceeds a third predefined threshold (which is preferably 0.66) (affirmative result in step **222**), then the current frame is selected as a key-frame for that motion activity segment, all the cumulative sums are reset to zero, and the

The cited section above contains no mention of standard deviations or motion vector magnitudes. Toklu does not contemplate motion vectors as they are only available in the compressed domain and, therefore, cannot teach the claimed measuring.

Regarding claim 9, claimed is partitioning the compressed video into a plurality of segments according to shot boundaries, measuring the cumulative motion activity intensity in each segment, selecting the key frames from each segment according to the cumulative motion activity intensity and concatenating the key-frames in a temporal order to form a summary of the compressed video. The Examiner cites Toklu, column 5, lines 40-45, as teaching the first element of the claim, i.e., the claimed partitioning. However, the Examiner fails to provide support for his rejection of any of the other elements of the claim. The Applicant, therefore, regards

claim 9 as allowable.

Regarding claim 10, claimed is partitioning the compressed video into a plurality of segments using a binary partitioning based on the cumulative motion activity intensity, wherein a first partitioning generates one segment that includes the entire video, a second partitioning generates two segments, each segment of the second partition having a substantially equal amount of cumulative motion activity intensity, and each subsequent partitioning dividing segments of a previous partitioning into two equal halves based on the accumulated motion activity intensity until a predetermined number of segments are generated, selecting a middle frame from each segment according to the cumulative motion activity intensity as one of the key-frames and concatenating the key-frames in a temporal order to form a summary of the compressed video. The Examiner cites Toklu, column 5, lines 50-55, as teaching the first element of the claim, i.e., the claimed binary partitioning. However, the cited section (reproduced below) makes no mention of binary partitioning.

> issued on Nov. 10, 1998, entitled "Apparatus For Detecting A Cut In A Video," which is commonly owned and incorporated herein by reference. The above-incorporated cut detection method partitions video data into a set of "shots" comprising visually abrupt cuts or camera breaks. As stated above, a video "shot" represents a contiguous recording of

The cut detection method of Toklu partitions the video based on visual cuts or camera breaks without regarding to the motion content of the segments produced by the partitioning. Also, the Examiner cites Toklu, column 9, lines 55-60 (reproduced below), as teaching the second element of the claim, i.e., that the claimed partitioning is based on the cumulative motion activity intensity.

> median filtered using a 7-tab filter (as is understood by those skilled in the art). The next step is to eliminate holes that are less than or equal to 15 frames long and finally, to eliminate the hills that are less than or equal to 15 frames long. The groups of frames of the binary MA curve that remain after the smoothing process are deemed motion activity segments

There is no mention of partitioning in this section. Instead, it describes the curve smoothing process of Toklu, which cannot teach the claimed partitioning based on cumulative motion activity intensity. Finally, the Examiner fails to provide support for his rejection of any of the other elements of the claim. The Applicant, therefore, regards claim 9 as allowable.

Regarding independent claim 11, claimed is selecting key-frames in a progressive ordering from the compressed video until a termination condition is reached and concatenating the key-frames in a temporal order to form the progressive summary of the compressed video. The progressive ordering is done by selecting a first frame of the compressed video as a first key-frame, selecting a last frame of the compressed video as a second frame, measuring a cumulative motion activity intensity in the compressed video, selecting a middle frame from the compressed video according to the cumulative motion activity intensity as a third key frame, partitioning the compressed video into two equal segments according to the motion activity intensity, selecting a middle frame according to the cumulative motion activity intensity of each segment as a fourth and fifth key frame, iteratively partitioning each previously partitioned segment into two smaller equal sized segments according to the motion activity intensity, and selecting further middle frames according to the cumulative motion activity intensity from each smaller sized segment as two next key-frames until the termination condition is reached.

The Examiner cites Toklu, column 6, lines 42-60, as teaching an element of the claim, namely, the selecting of key-frames in a progressive ordering from the compressed video until a termination condition is reached. The cited section is reproduced below.

> An exemplary method of operation of the system of FIG. **1** according to the present invention will now be explained in detail with reference to the flow diagrams of FIGS. **2A**, **2B** and **2C**. More specifically, the flow diagrams of FIGS. **2A**, **2B** and **2C** illustrate various key-frame selection processes according to different levels of operation (e.g., levels A, B, C and D as noted above). It is to be understood that the following description of the combination of operating levels and key-frame methods are for purposes of illustration and that other combinations of key-frame selection methods and operation levels may be envisioned by those skilled in the art.
>
> Referring to FIG. **2A**, video data (e.g., a video/multimedia file, video data stream, etc.) is input to the system **10** (step **200**) and then partitioned into a plurality of video segments (step **201**). As stated above, the video data is preferably segmented into video shots using the cut detection method described in the above-incorporated U.S. Pat. No. 5,835, 163.

As was shown above, Toklu selects key-frames using threshold criteria, not according to cumulative motion activity intensity, as claimed. Further, Toklu does not contemplate motion vectors and cannot teach selecting key-frames according to cumulative motion activity intensity, as claimed.

The Examiner cites Toklu, column 6, lines 35-41, as teaching an element of the claim, namely, the concatenating the key-frames in a temporal order to form the progressive summary of the compressed video. The cited section is reproduced below.

> The output of the key-frame elimination module **18** is list of key-frames, e.g., keyframelist, for each video shot. The key-frame list may be stored in a key-frame list database **19**. The key-frame lists may be used by video processing/ indexing systems as components for, e.g., multimedia and/or video summarization.

As above, Toklu makes no mention of a temporal order for the list of key-frames. Further, the invention concatenates key-frames in compressed video. Toklu's pixel-based method does not operate in the compressed domain and cannot teach the claimed concatenating.

The Examiner cites Toklu, column 7, lines 10-15, as teaching an element of the claim, namely, the selecting of a first frame of the compressed video as a first key-frame. The cited section is reproduced below.

$$PD(t) = \frac{1}{N} \sum_{(x,y) \in S} |I(t, x, y) - I(t-1, x, y)|, \quad t_f \geq t \geq t_i \qquad \text{Eqn. 1}$$

where t_i and t_f are the starting and ending frame number, respectively, of a given video segment (determined from the segment boundary data), N denotes the total number of

As was shown above, Toklu partitions uncompressed video into "shots" bounded by "visually abrupt cuts or camera breaks," not segments having a substantially equal cumulative motion activity as determined by measuring average motion vector magnitude in compressed video, as claimed. Also, the method of Toklu selects key-frames using threshold criteria, not according to cumulative motion activity intensity, as claimed. Furthermore, Applicant can find no specific rationale in the Examiner's action for the rejection of this element in the above section. The Applicant respectfully requests that the Examiner specifically point out which parts of the above section teach the selecting of a first frame of the compressed video as a first key-frame.

The Examiner again cites Toklu, column 7, lines 10-15, as teaching an element of the claim, namely, the selecting of a last frame of the compressed video as a second key-frame. As was shown above, Toklu partitions uncompressed video into "shots" bounded by "visually abrupt cuts or camera breaks," not segments having a substantially equal cumulative motion activity as determined by measuring average motion vector magnitude in compressed video, as claimed. Also, the method of Toklu selects key-frames using threshold criteria, not according to cumulative motion activity intensity, as claimed. Furthermore, Applicant can find no specific rationale in the Examiner's action for the rejection of this element in the cited section. The Applicant respectfully requests that the Examiner specifically point out which parts of the above section teach the selecting of a last frame of the compressed video as a second key-frame.

The Examiner cites Toklu, column 9, lines 17-46, as teaching an element of the claim, namely, the measuring of a cumulative motion activity intensity in the compressed video. As was shown above, Toklu employs a pixel-based method operating on uncompressed video. Toklu estimates camera motion

from dissimilarity measures obtained by comparing images in uncompressed video on a frame-by-frame basis. Toklu cannot teach the claimed measuring of a cumulative motion activity intensity using average motion vector magnitudes, which are only available in the compressed domain.

The Examiner cites Toklu, column 12, lines 50-65, as teaching an element of the claim, namely, the selecting of a middle frame from the compressed video according to the cumulative motion activity intensity as a third key frame. The cited section is reproduced below.

> performed one at a time (in any order) or in parallel.
>
> After key-frame selection process is complete, the final step (as indicated above) is to eliminate visually similar key-frames (step **210**, FIG. 2A). An elimination process in accordance with a preferred embodiment is based on based on histogram or perceptual color similarity (as discussed below). It is to be understood that key-frame selection processes described above may not generate visually different images in every instance. In some situations, several key-frames may not comprise "new" information, notwithstanding that there is significant motion activity or histogram activity among such key-frames. Furthermore, since an operating level (e.g., level D discussed above) may utilize a plurality of key-frame selection processes, the probability of having an overlap of selected key-frames (from the multiple processes) being visually similar increases.

Applicant fails to see where in this section is taught the selecting of a key-frame. The cited section describes an elimination process to remove visually similar key-frames. Toklu fails to teach the claimed limitation of selecting a middle frame from the compressed video according to the cumulative motion activity intensity as a third key frame.

The Examiner cites Toklu, column 5, lines 50-60, as teaching an element of the claim, namely, the partitioning of the compressed video. The cited section is reproduced below.

> issued on Nov. 10, 1998, entitled "Apparatus For Detecting A Cut In A Video," which is commonly owned and incorporated herein by reference. The above-incorporated cut detection method partitions video data into a set of "shots" comprising visually abrupt cuts or camera breaks. As stated above, a video "shot" represents a contiguous recording of one or more video frames depicting a continuous action in time and space. The above-incorporated cut detection method segments video data into shots using an integrated pixel-based difference metric (discussed below) and distribution-based difference metric.

Again, the method of Toklu can never anticipate the claimed partitioning because it operates on uncompressed video and creates "shots" bounded by "visually abrupt cuts or camera breaks," not segments having a substantially equal cumulative motion activity as determined by measuring average motion vector magnitude in compressed video.

The Examiner cites Toklu, column 10, lines 13-20, as teaching an element of the claim, namely, partitioning the compressed video into two equal segments. The cited section is reproduced below.

> have the first frame as the selected key-frame.
>
> On the other hand, if one or more motion activity segments were detected for a given video segment (affirmative result in step **216**), then the next step, in general, is selecting key-frames for a given shot by sampling the MA curve (non-uniformly in time). More specifically, in a preferred embodiment, key-frames for a given shot are selected from each of the detected motion activity segments for the given

Applicant fails to see where in this section is taught the partitioning of uncompressed video into two equal segments. The cited section describes the selecting of key-frames. Toklu fails to teach the claimed limitation of partitioning the compressed video into two equal segments.

The Examiner cites Toklu, column 9, lines 45-55, as teaching an element of the claim, namely, partitioning the compressed video into two equal segments according to the motion activity intensity. The cited section is reproduced below.

> median filtered using a 7-tab filter (as is understood by those skilled in the art). The next step is to eliminate holes that are less than or equal to 15 frames long and finally, to eliminate the hills that are less than or equal to 15 frames long. The groups of frames of the binary MA curve that remain after the smoothing process are deemed motion activity segments within the shot. In particular, the remaining hills in the activity curve become the activity segments, i.e., the values of the binary motion activity curve for a number of consecutive frames (at least 15 frames) are equal to "1."
>
> It is to be understood that the selected frame lengths in the

Applicant fails to see where in this section is taught the partitioning the compressed video into two equal segments according to the motion activity intensity. The cited section describes the sampling of a curve to produce segments using predetermined criteria, which would not result in a partitioned video of two equal segments. Toklu fails to teach the claimed limitation of partitioning the compressed video into two equal segments according to the motion activity intensity.

The Examiner cites Toklu, column 12, lines 50-65, as teaching an element of the claim, namely, selecting a middle frame according to the cumulative motion activity intensity of each segment as a fourth and fifth key frame. The cited section is reproduced below.

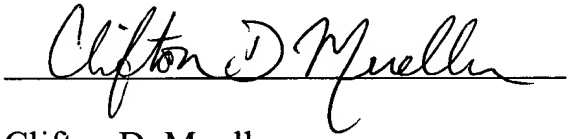
> performed one at a time (in any order) or in parallel.
>
> After key-frame selection process is complete, the final step (as indicated above) is to eliminate visually similar key-frames (step **210**, FIG. **2A**). An elimination process in accordance with a preferred embodiment is based on based on histogram or perceptual color similarity (as discussed below). It is to be understood that key-frame selection processes described above may not generate visually different images in every instance. In some situations, several key-frames may not comprise "new" information, notwithstanding that there is significant motion activity or histogram activity among such key-frames. Furthermore, since an operating level (e.g., level D discussed above) may utilize a plurality of key-frame selection processes, the probability of having an overlap of selected key-frames (from the multiple processes) being visually similar increases.

Applicant fails to see where in this section is taught the selecting of a key-frame. The cited section describes an elimination process to remove visually similar key-frames. Toklu fails to teach the claimed limitation of selecting a middle frame according to the cumulative motion activity intensity of each segment as a fourth and fifth key frame.

The Examiner cites Toklu, column 10, lines 30-50, as teaching an element of the claim, namely, iteratively partitioning each previously partitioned segment into two smaller equal sized segments according to the motion activity intensity, and selecting further middle frames according to the cumulative motion activity intensity from each smaller sized segment as two next key-frames until the termination condition is reached. As was shown above, Toklu can never teach the claimed partitioning and selecting.

It is believed that this application is now in condition for allowance. A notice to this effect is respectfully requested. Should further questions arise concerning this application, the Examiner is invited to call Applicant's agent at the number listed below. Please charge any shortage in fees due in connection with the filing of this paper to Deposit Account 50-0749.

Respectfully submitted,
Mitsubishi Electric Research Laboratories, Inc.

By 

Clifton D. Mueller
Agent for the Assignee
Reg. No. 57,836

201 Broadway, 8th Floor
Cambridge, MA 02139
Telephone: (617) 621-7517
Customer No. 022199